



Revista Hambatu Science

Ambato – Ecuador / ISSN 3151-815X (en línea) / julio - septiembre 2026

Volumen 1, Número 3

<https://doi.org/10.63862/rhs-v1n3-746-769-2026>

Modelo longitudinal para la identificación temprana del riesgo de deterioro sostenido de la función renal en pacientes de programas de riesgo cardiovascular del Caribe colombiano

*Longitudinal model for early identification of the risk of
sustained renal function decline in patients enrolled in
cardiovascular risk programs in the Colombian Caribbean*

Jair de Jesús Díaz Lora

Investigador independiente en analítica de datos aplicada en salud



DOI : <https://doi.org/10.63862/rhs-v1n3-746-769-2026>

Modelo longitudinal para la identificación temprana del riesgo de deterioro sostenido de la función renal en pacientes de programas de riesgo cardiovascular del Caribe colombiano

Jair de Jesús Díaz Lora

Investigador independiente en analítica de
datos aplicada en salud

jairdiazlora@gmail.com

<https://orcid.org/0009-0006-0871-5096>

Barranquilla, Colombia

Recibido: 2026-14-09

Aceptado: 2026-15-09

Publicado: 2026-17-09

Resumen

Las enfermedades cardiovasculares, la hipertensión arterial, la diabetes mellitus y la enfermedad renal crónica conforman un entramado clínico en el que el deterioro de la función renal puede progresar silenciosamente. Esta segunda investigación de la línea longitudinal del autor analiza la progresión del deterioro funcional renal, entendido operacionalmente como una reducción sostenida de la tasa de filtración glomerular (TFG), sin constituir una nueva definición diagnóstica. Se realizó un estudio observacional retrospectivo, cuantitativo, longitudinal y predictivo sobre 191.096 registros correspondientes a 26.717 pacientes únicos. La arquitectura PALIC-PY-10 V12.1.0 construyó 18.326 landmarks, uno por paciente, con horizonte prospectivo de 365 días; 15.430 fueron evaluables y se registraron 124 eventos primarios (0,804%). El desenlace exigió una reducción sostenida $\geq 40\%$ de la TFG respecto del valor índice, con cruce a $\leq 60\%$ y confirmación al menos 30 días después. En la prueba temporal canónica, Random Forest alcanzó ROC-AUC 0,8689, PR-AUC 0,4956 y Brier 0,00619. El bootstrap de 1.500 remuestras obtuvo IC95% de 0,7912–0,9436 para ROC-AUC y 0,3433–0,6645 para PR-AUC. PALIC-PY-10 constituye además un framework metodológico-computacional desarrollado y fortalecido mediante seis auditorías científicas adversariales. Los resultados representan evidencia interna prometedora, pero no constituyen validación externa, utilidad clínica, diagnóstico ni demostración de generalización fuera del contexto estudiado.

Palabras clave: enfermedad renal crónica, tasa de filtración glomerular, aprendizaje automático, análisis longitudinal, riesgo cardiovascular, Random Forest, Caribe colombiano.

Longitudinal model for early identification of the risk of sustained renal function decline in patients enrolled in cardiovascular risk programs in the Colombian Caribbean

Abstract

Cardiovascular diseases, arterial hypertension, diabetes mellitus, and chronic kidney disease form a complex clinical network in which renal function deterioration may progress silently. This study examines the progression of renal functional decline, operationally defined as a sustained reduction in estimated glomerular filtration rate (eGFR), without proposing a new diagnostic definition. A retrospective, quantitative, longitudinal, and predictive observational study was conducted using 191,096 records from 26,717 unique patients. The PALIC-PY-10 V12.1.0 architecture generated 18,326 landmarks, one per patient, with a 365-day prospective horizon; 15,430 were evaluable, and 124 primary events (0.804%) were observed. The outcome required a sustained $\geq 40\%$ reduction in eGFR from the index value, crossing to $\leq 60\%$, with confirmation at least 30 days later. In the canonical temporal test, Random Forest achieved a ROC-AUC of 0.8689, PR-AUC of 0.4956, and Brier score of 0.00619. Bootstrap analysis with 1,500 resamples yielded 95% CIs of 0.7912–0.9436 for ROC-AUC and 0.3433–0.6645 for PR-AUC. PALIC-PY-10 also represents a methodological-computational framework developed and strengthened through six adversarial scientific audits. These findings provide promising internal evidence but do not constitute external validation, clinical utility, diagnosis, or evidence of generalizability beyond the studied context.

Keywords: chronic kidney disease, estimated glomerular filtration rate, machine learning, longitudinal analysis, cardiovascular risk, Random Forest, Colombian Caribbean.

Introducción

En Colombia, la ERC coexiste con hipertensión arterial y diabetes mellitus en una población con elevada carga cardiovascular y asistencial. Los informes nacionales de la Cuenta de Alto Costo han mostrado la magnitud de estas condiciones y la necesidad de fortalecer la gestión basada en información clínica y seguimiento poblacional (Cuenta de Alto Costo, 2024). En programas de riesgo cardiovascular, la función renal no constituye un desenlace aislado: interactúa con hipertensión, diabetes, edad, tratamiento, continuidad del cuidado y disponibilidad de controles de laboratorio. Por ello, la oportunidad para anticipar un deterioro renal relevante depende tanto del algoritmo como de la calidad temporal de la información que alimenta el modelo.

La literatura reciente demuestra que el aprendizaje automático puede apoyar la predicción de progresión renal. Modelos como KidneyIntelX han combinado información clínica y biomarcadores mediante Random Forest para identificar riesgo de deterioro renal, incluyendo desenlaces que incorporan una disminución sostenida del 40% de la TFG o insuficiencia renal (Chan et al., 2021). Otros modelos han utilizado Random Forest con grandes conjuntos de variables de laboratorio y han logrado validación externa para desenlaces de caída $\geq 40\%$ de la TFG o falla renal (Tangri et al., 2022). La validación de Klinrisk en poblaciones de ensayos clínicos también mostró capacidad discriminativa y calibración para predecir progresión de ERC hasta tres años (Tangri et al., 2024).

En Colombia existe además un antecedente especialmente próximo. Rojas et al. (2025) desarrollaron modelos interpretables basados en XGBoost para predecir la TFG futura a seis, nueve y doce meses en más de 29.000 pacientes con ERC del Caribe colombiano, con validación interna y externa y uso de SHAP. Este antecedente confirma la pertinencia del aprendizaje automático en la región, pero aborda una pregunta distinta: la predicción de un valor futuro de TFG en pacientes con ERC, mientras el presente estudio se orienta a la identificación de un evento binario operacional de deterioro sostenido $\geq 40\%$ en pacientes pertenecientes a programas de riesgo cardiovascular. En consecuencia, ambos trabajos son complementarios y no corresponden a la misma investigación ni utilizan la misma cohorte analítica del presente manuscrito.

El problema metodológico central de los modelos longitudinales no es únicamente alcanzar un alto AUC. La construcción del conjunto de predictores debe respetar la disponibilidad temporal de la información, evitar que variables futuras entren al modelo y distinguir claramente la fecha índice del periodo en el que se observa el desenlace. La literatura metodológica ha advertido que los estudios de predicción clínica basados en aprendizaje automático presentan con frecuencia problemas de sesgo, sobreajuste, manejo de datos faltantes y validación inadecuada (Moons et al., 2019; Collins et al., 2024). La reciente guía TRIPOD+AI recomienda reportar de manera transparente el desarrollo y evaluación de modelos predictivos basados tanto en regresión como en aprendizaje automático.

El primer artículo de esta línea de investigación analizó el riesgo de abandono asistencial y deterioro clínico en programas de riesgo cardiovascular del Caribe colombiano (Díaz Lora, 2026). El presente estudio mantiene la lógica poblacional y longitudinal, pero desplaza la pregunta hacia la trayectoria renal y construye un desenlace temporalmente definido de deterioro sostenido.

Tabla 1.
Sinergia entre el Artículo 1 y el Artículo 2

Dimensión	Artículo 1	Artículo 2
Pregunta	Abandono, continuidad y deterioro clínico en programas de riesgo cardiovascular.	Deterioro sostenido de la función renal dentro de 365 días.
Población	26.732 pacientes de Atlántico y Cesar.	Fuente integrada de 26.717 pacientes; cohorte renal y landmarks definidos para la segunda pregunta.
Señal principal	Continuidad, adherencia, edad y control de HTA.	Trayectoria renal histórica, TFG/creatinina y variables clínicas disponibles al índice.
Modelo	Random Forest + SHAP.	Random Forest + regresión logística comparadora, evaluación temporal y bootstrap.
Resultado	AUC $\approx$ 0,659.	ROC-AUC 0,8689; PR-AUC 0,4956; Brier 0,00619.
Contribución	Primera investigación longitudinal de la línea sobre continuidad y deterioro clínico.	Segunda investigación: progresión funcional renal operacionalizada y PALIC-PY-10 V12.1.0 endurecido.

La contribución metodológica central se articula alrededor de PALIC, Plataforma Analítica Longitudinal para Investigación Clínica. Para este estudio se utilizó el módulo PALIC-PY-10

V12.1.0, concebido como una infraestructura metodológica y computacional versionada para validar la fuente, construir cohortes longitudinales, definir landmarks, establecer el desenlace, generar predictores point-in-time, controlar leakage, modelar, evaluar robustez y conservar trazabilidad reproducible. El framework es propio del autor en cuanto a su arquitectura y metodología; Random Forest, regresión logística, Python y las técnicas estadísticas empleadas son herramientas y métodos existentes, no invenciones atribuidas al autor.

Objetivo general

Desarrollar y evaluar un modelo longitudinal para la identificación temprana del riesgo de deterioro sostenido de la función renal, definido operacionalmente como una disminución $\geq 40\%$ de la TFG respecto de su valor índice y confirmada temporalmente, en pacientes pertenecientes a programas de riesgo cardiovascular del Caribe colombiano.

Objetivos específicos

Construir una cohorte renal longitudinal y un landmark por paciente que permita aplicar un horizonte de predicción de 365 días y clasificar explícitamente la evaluabilidad según la disponibilidad de observaciones futuras.

Definir un desenlace operacional de deterioro sostenido de la TFG y evaluar sensibilidades relacionadas con magnitud de descenso y tiempo de confirmación.

Desarrollar un conjunto de predictores estrictamente point-in-time, excluyendo variables futuras y variables agregadas del Artículo 1.

Evaluar Random Forest y regresión logística penalizada mediante validación temporal, ROC-AUC, PR-AUC, Brier, calibración y bootstrap.

Examinar la estabilidad del Random Forest mediante cinco semillas y explorar la contribución relativa de las variables mediante importancia nativa y permutación.

Identificar limitaciones de calidad de datos, representatividad, frecuencia del evento y necesidad de validación externa antes de cualquier uso clínico.

Estado del arte

La evidencia clínica contemporánea considera la TFG y la albuminuria como componentes centrales para caracterizar el riesgo renal. KDIGO 2024 incorpora modelos de riesgo y destaca que la progresión puede presentarse en diferentes categorías de función renal; asimismo, reconoce la relevancia de desenlaces que incluyen una disminución sustancial de la TFG. En este contexto, un desenlace basado en caída sostenida de 40% resulta metodológicamente conectado con una tradición de investigación clínica y de modelamiento predictivo, aunque la definición exacta utilizada en este estudio es operacional y no constituye por sí sola una nueva definición diagnóstica de ERC.

Los modelos de aprendizaje automático han ampliado la capacidad para integrar múltiples mediciones y variables clínicas. El estudio de Tangri et al. (2022) utilizó Random Forest y datos de laboratorio para predecir progresión de ERC definida por caída de 40% de la TFG o falla renal, con AUC de 0,88 a dos años en validación interna y 0,87 en validación externa. La similitud del desenlace permite ubicar el presente estudio dentro de una línea metodológica consolidada, pero la cohorte, el contexto asistencial, el horizonte de 365 días y la arquitectura temporal son diferentes.

KidneyIntelX representa otro antecedente relevante porque combinó datos electrónicos y biomarcadores mediante Random Forest y evaluó un desenlace compuesto que incluía disminución sostenida de 40% de la TFG. La importancia de este trabajo para el presente estudio no radica en reproducir su modelo, sino en mostrar que la predicción de deterioro renal requiere integrar información longitudinal y evaluar el comportamiento del modelo frente a desenlaces poco frecuentes. En el presente estudio, la frecuencia del evento primario fue 0,804%, lo que obliga a interpretar el PR-AUC y la calibración con mayor cautela que un AUC aislado.

Klinrisk constituye otro punto de comparación. En su validación externa en CANVAS/CREDENCE, el modelo logró AUC de 0,81 a un año y 0,88 a tres años para un desenlace de $\geq 40\%$ de disminución de TFG o falla renal, acompañado de evaluación de Brier y calibración. Este antecedente refuerza la necesidad de reportar métricas complementarias al AUC. Sin embargo, Klinrisk se desarrolló en poblaciones clínicas y marcos temporales

distintos, mientras que el presente modelo fue construido sobre datos observacionales reales de programas de riesgo cardiovascular del Caribe colombiano y con validación temporal interna.

El antecedente colombiano de Rojas et al. (2025) es particularmente importante para establecer el vacío regional. El estudio incluyó más de 29.000 pacientes con ERC de la región Caribe y desarrolló modelos XGBoost para estimar la siguiente TFG a 6, 9 y 12 meses, con validación interna y externa y explicabilidad mediante SHAP. El presente trabajo no pretende sustituir ni duplicar ese enfoque. Su diferencia se encuentra en la unidad de predicción: no se estima directamente el próximo valor de TFG, sino la probabilidad de cumplir un evento sostenido de deterioro $\geq 40\%$ durante un horizonte de 365 días, en una población de programas de riesgo cardiovascular y mediante una arquitectura longitudinal point-in-time con auditoría explícita de leakage.

Desde el punto de vista de reporte, TRIPOD+AI plantea que los estudios de predicción deben describir con precisión población, desenlace, predictores, partición de datos, desarrollo del modelo, evaluación, validación y prácticas de ciencia abierta (Collins et al., 2024). PROBAST complementa este marco al evaluar calidad, aplicabilidad y riesgo de sesgo (Moons et al., 2019). Estas recomendaciones son especialmente pertinentes en estudios con eventos infrecuentes, pues la aparente precisión de un modelo puede ocultar inestabilidad, sobreajuste o mala calibración.

En conjunto, la literatura permite afirmar que el presente estudio no constituye el primer modelo de aprendizaje automático para progresión renal, ni el primer trabajo predictivo realizado en pacientes del Caribe colombiano. La novedad defendible se encuentra en la combinación específica de: población vinculada a programas de riesgo cardiovascular; trayectoria renal de datos reales; desenlace operacional de deterioro sostenido $\geq 40\%$; horizonte de 365 días; landmark único por paciente; validación temporal 70/30; control point-in-time y de leakage; y utilización de PALIC como infraestructura metodológica versionada y auditable. La búsqueda realizada no identificó un artículo idéntico que reúna simultáneamente estas características, aunque sí identificó antecedentes muy próximos que deben reconocerse expresamente.

Metodología

Diseño y enfoque

Se realizó un estudio observacional retrospectivo, cuantitativo, longitudinal y predictivo. El análisis se desarrolló sobre datos secundarios de atención en salud previamente integrados. El propósito fue construir y evaluar internamente un modelo de riesgo, no establecer diagnóstico individual ni demostrar utilidad clínica.

Fuente de información y trazabilidad

La fuente analítica contenía 191.096 registros, 53 columnas y 26.717 pacientes únicos. No se identificaron duplicados completos y se registraron 3.930.280 valores nulos en el conjunto fuente. Se detectaron 36 registros con TFG ausente, sin valores no positivos. La ventana de observaciones renales válidas comprendió del 1 de abril de 2024 al 31 de marzo de 2026. La fuente quedó identificada mediante :

SHA-256: 8228023f37e3f25d3695d7e588ed0704abd0840d9ffa6019432f2b746d76af3e.

Manejo de datos faltantes. La fuente presentó 3.930.280 valores nulos. La construcción de la cohorte renal descartó únicamente registros sin identificadores o fecha de laboratorio utilizables y observaciones sin TFG válida cuando eran necesarias para definir la trayectoria renal. En el modelamiento, las variables numéricas con valores faltantes fueron imputadas mediante la mediana y las variables categóricas mediante la categoría más frecuente, utilizando parámetros aprendidos exclusivamente en el conjunto de entrenamiento; las variables que no aportaron ningún valor observado en entrenamiento fueron retiradas de la matriz efectiva. La antropometría continua quedó sin valores utilizables después de los filtros y se conservó únicamente la bandera AS-OF de disponibilidad. Esta estrategia evita que la información del conjunto de prueba intervenga en el aprendizaje del tratamiento de faltantes.

PALIC: Plataforma Analítica Longitudinal para Investigación Clínica

El procesamiento, construcción de la cohorte y modelamiento se desarrollaron mediante PALIC (Plataforma Analítica Longitudinal para Investigación Clínica), un framework metodológico y computacional propio orientado al análisis longitudinal de datos de salud. Para el presente

estudio se utilizó el módulo PALIC-PY-10 V12.1.0, que integra validación y trazabilidad de la fuente, construcción de cohortes longitudinales y landmarks, definición operacional del desenlace, generación de predictores point-in-time, control explícito de información futura, modelamiento, evaluación, análisis de robustez, sensibilidad y auditoría reproducible.

La versión V12.1.0 establece reglas explícitas para la temporalidad de los predictores, la construcción del landmark, la definición operacional del desenlace, la prevención de leakage y la trazabilidad de los resultados. La arquitectura final incorpora RowID estable, split temporal canónico persistido, predicciones trazables, hashes de fuente y código, manifiesto del entorno y autoverificación final. PALIC no se presenta como un algoritmo nuevo: su aporte propio es la arquitectura metodológica que gobierna y hace auditable el ciclo completo del análisis.

Cohorte longitudinal y landmark

De la fuente integrada se obtuvieron 189.349 observaciones renales correspondientes a 25.001 pacientes con trayectoria longitudinal. V12.1.0 seleccionó un único landmark por paciente: el último punto elegible situado al menos 365 días antes del máximo administrativo de la fuente y con al menos una TFG válida previa. La regla de selección del índice no utilizó observaciones futuras del paciente. Se obtuvieron 18.326 landmarks, con fechas índice entre el 11 de abril de 2024 y el 31 de marzo de 2025.

Tabla 2.

Construcción de la cohorte longitudinal y de los landmarks

Etap	Pacientes/observaciones	Criterio
Fuente analítica	26.717 pacientes / 191.096 registros	Fuente integrada y trazable
Cohorte renal	25.001 pacientes / 189.349 observaciones	TFG y creatinina válidas; trayectoria longitudinal
Pacientes con ≥ 2 fechas	25.001	Al menos dos fechas renales distintas
Landmarks	18.326	Último landmark elegible antes del máximo administrativo -365 días; ≥ 1 TFG previa
Evaluable	15.430 (84,20%)	≥ 2 TFG futuras válidas dentro de 365 días
No evaluables	2.896 (15,80%)	Censurados/no evaluables; nunca recodificados como no evento
Eventos primarios	124 (0,804%)	Deterioro sostenido $\geq 40\%$

Definición operacional del desenlace

El desenlace primario Target_Sustained_40 requirió: (1) una TFG futura $\leq 60\%$ de la TFG índice, equivalente a una caída de al menos 40%; (2) una segunda medición posterior que confirmara el cruce; y (3) una separación mínima de 30 días entre ambas mediciones, dentro de los 365 días posteriores al índice. Las sensibilidades consideraron umbrales de $\geq 30\%$ y $\geq 50\%$, y confirmaciones a 30, 60 y 90 días. La insuficiencia de observación futura se trató como no evaluabilidad y nunca como ausencia de evento.

Tabla 3.
Definición operacional del desenlace

Definición	Valor
Desenlace primario	Target_Sustained_40
Magnitud del descenso	$\geq 40\%$ respecto de TFG índice
Cruce del umbral	TFG futura $\leq 60\%$ de TFG índice
Confirmación	Segunda medición bajo umbral ≥ 30 días después
Horizonte	365 días
Sensibilidades de magnitud	$\geq 30\%$ y $\geq 50\%$
Sensibilidad de confirmación	30, 60 y 90 días
Desenlaces exploratorios	Empeoramiento de categoría G y pendiente futura de TFG

Predictores y principio point-in-time

Los predictores fueron contruidos exclusivamente con información disponible hasta la fecha índice. La matriz efectiva de 28 variables incluyó sexo, región, patología, edad, departamento, TFG y creatinina índice y previas, diferencias entre mediciones, pendientes históricas, variabilidad de TFG, recencia, frecuencia y cantidad de mediciones, además de una bandera de disponibilidad antropométrica AS-OF. Las familias futuras del desenlace y las variables agregadas del Artículo 1 fueron excluidas del modelamiento.

La antropometría se trató con una estrategia temporal AS-OF. La cobertura efectiva de disponibilidad antropométrica fue baja y las variables continuas antropométricas no aportaron valores utilizables en la matriz final después de los filtros de entrenamiento. La decisión fue conservar la bandera de disponibilidad como indicador de proceso, sin utilizar información GLOBAL-FIRST ni información posterior al landmark.

Modelos predictivos

Se desarrollaron dos familias principales: Random Forest como modelo de referencia no lineal (Breiman, 2001) y regresión logística penalizada como modelo comparador. El Random Forest se configuró con 350 árboles, `balanced_subsample`, `max_features=sqrt`, `min_samples_leaf=5` y semilla primaria 42, con `n_jobs=1` para estabilidad de ejecución. La regresión logística utilizó `class_weight=balanced` y `C=0,5`.

Tabla 4.
Configuración de los modelos y esquema de evaluación

Componente	Configuración
Modelo principal	Random Forest
Árboles	350
Balanceo	<code>balanced_subsample</code>
<code>max_features</code>	<code>sqrt</code>
<code>min_samples_leaf</code>	5
Modelo comparador	Regresión logística penalizada
<code>class_weight</code>	<code>balanced</code>
C	0,5
Partición	Temporal 70/30 por FechaÍndice
Random split	Prohibido
Bootstrap	1.500 iteraciones
Semillas de robustez	7, 21, 42, 77 y 101

El desempeño se evaluó mediante ROC-AUC, PR-AUC y Brier score. La calibración se examinó mediante intercepto y pendiente. La partición principal fue temporal, con fecha de corte canónica 2025-03-28, 10.918 landmarks de entrenamiento y 4.512 de prueba, sin fechas índice compartidas. Se calcularon bootstrap de 1.500 remuestras para el modelo primario, robustez en cinco semillas, análisis Late300, observabilidad y subgrupos exploratorios.

Justificación de la validación temporal. La evaluación principal se realizó mediante una partición temporal porque el objetivo es estimar el comportamiento del modelo frente a información cronológicamente posterior al periodo utilizado para su entrenamiento. Una validación cruzada aleatoria mezclaría observaciones de distintos momentos y produciría una estimación menos representativa del escenario prospectivo que motiva el estudio. Por ello, la validación temporal 70/30 se definió como análisis principal; el bootstrap y las cinco semillas se utilizaron como procedimientos complementarios para cuantificar incertidumbre y estabilidad, no como sustitutos de la separación temporal.

Debido a la prevalencia extremadamente baja del evento, la interpretación se centró en un conjunto de métricas complementarias. ROC-AUC resume discriminación global; PR-AUC es especialmente informativa cuando los eventos son infrecuentes; Brier evalúa el error probabilístico; y la calibración permite examinar la correspondencia entre probabilidad predicha y frecuencia observada. Ninguna de estas métricas, por sí sola, demuestra utilidad clínica.

La PR-AUC recibió especial atención porque el evento primario ocurrió en 0,804% de los casos evaluables. En este contexto, un ROC-AUC elevado puede coexistir con valores predictivos positivos modestos. Por ello, la interpretación final se realizó conjuntamente con PR-AUC, Brier, calibración, incertidumbre bootstrap y sensibilidad entre semillas.

Auditoría de leakage y control de calidad

PALIC-PY-10 V12.1.0 ejecutó un gate de leakage que verificó ausencia de variables futuras, exclusión de variables agregadas del Artículo 1, ausencia de GLOBAL-FIRST como predictor, exclusión de familias Target_, CrossDate_, ConfirmDate_, N_Future_, MinTFGFuture_, MaxDeclineFuture_, Slope_Annual_Future y stage_future, y cumplimiento estricto del principio point-in-time. El resultado final fue READY, con cero predictores prohibidos encontrados. El split temporal canónico fue persistido y verificado mediante RowID estable.

Consideraciones éticas

El análisis utilizó información secundaria procesada con fines analíticos, sin presentar identificadores personales en los resultados del manuscrito. La base fue tratada bajo criterios de confidencialidad y minimización de información. La documentación disponible para esta versión no contiene un número de acta, constancia de aprobación o resolución de exención emitida por un comité de ética que pueda ser verificada en el expediente analítico; por ello, el manuscrito no atribuye una aprobación o exención institucional que no esté documentada. La versión destinada al envío editorial deberá incorporar, cuando corresponda, la constancia institucional o declaración formal exigida por la revista. El estudio no se presenta como validación clínica ni como instrumento de decisión individual. La disponibilidad de datos de origen está restringida por contener información derivada de atención en salud.

Resultados

Caracterización de la cohorte

La fuente analítica comprendió 191.096 registros y 26.717 pacientes únicos. La cohorte renal válida quedó constituida por 189.349 observaciones y 25.001 pacientes con trayectoria longitudinal. V12.1.0 construyó 18.326 landmarks, uno por paciente. De ellos, 15.430 (84,20%) fueron evaluables para el horizonte de 365 días y 2.896 (15,80%) quedaron como no evaluables; estos últimos nunca fueron recodificados como no evento.

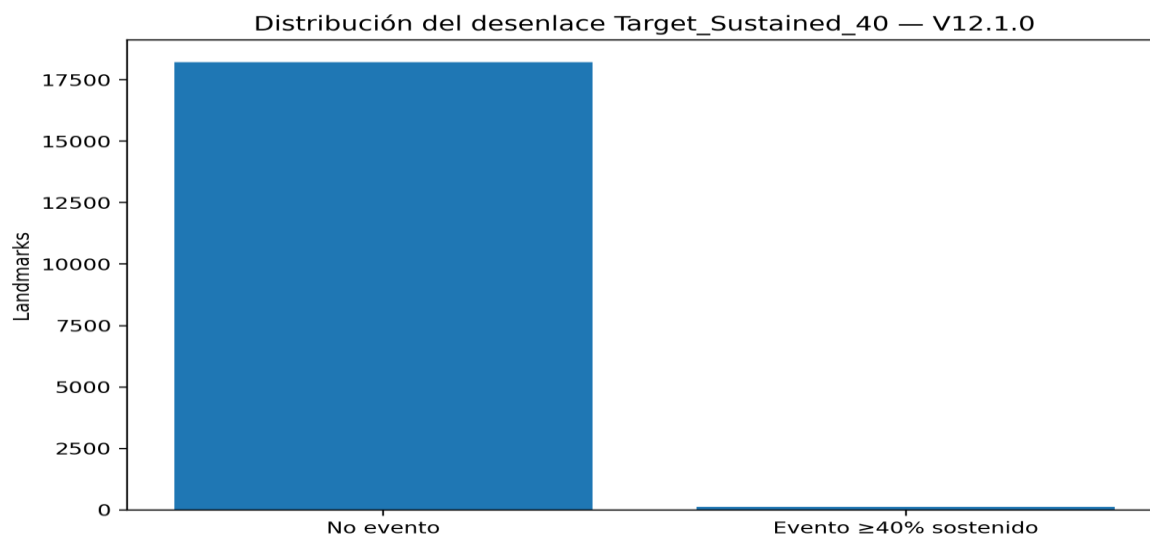
Tabla 5.

Caracterización de la cohorte final

Característica	Resultado
Registros fuente	191.096
Columnas	53
Pacientes únicos en fuente	26.717
Observaciones renales	189.349
Pacientes longitudinales	25.001
Fechas renales	2024-04-01 a 2026-03-31
Landmarks	18.326
Eventos $\geq 40\%$ sostenido	124
Frecuencia del evento	0,804% de evaluables
Evaluables 365 días	15.430 (84,20%)
No evaluables	2.896 (15,80%)

Figura 1.

Distribución del desenlace operacional Target_Sustained_40



Fuente: Elaboración propia mediante PALIC-PY-10 V12.1.0.

Sensibilidad del desenlace

Al modificar la magnitud del descenso, se observaron 249 eventos para el umbral de $\geq 30\%$ (1,614% de los evaluables) y 47 para $\geq 50\%$ (0,305%). Manteniendo la magnitud de 40% y ampliando la confirmación, se observaron 103 eventos con separación mínima de 60 días y 93 con separación mínima de 90 días. El desenlace primario de $\geq 40\%$ sostenido con confirmación ≥ 30 días comprendió 124 eventos (0,804%).

Tabla 6.
Sensibilidad del desenlace operacional

Escenario	Eventos	Frecuencia
$\geq 30\%$ de descenso	249	1,614%
$\geq 40\%$ sostenido, confirmación ≥ 30 días	124	0,804%
$\geq 50\%$ de descenso	47	0,305%
$\geq 40\%$, confirmación ≥ 60 días	103	—
$\geq 40\%$, confirmación ≥ 90 días	93	—

Desempeño predictivo

En la partición temporal canónica, el entrenamiento incluyó 10.918 landmarks y 84 eventos, mientras la prueba incluyó 4.512 landmarks y 40 eventos. El Random Forest de familia clínica completa obtuvo ROC-AUC 0,8689, PR-AUC 0,4956 y Brier 0,00619. La regresión logística obtuvo ROC-AUC 0,8421, PR-AUC 0,4680 y Brier 0,08190.

Tabla 7.
Desempeño predictivo en la partición temporal canónica

Modelo	ROC-AUC	PR-AUC	Brier	Eventos prueba
Random Forest — familia clínica completa	0,8689	0,4956	0,00619	40
Regresión logística — familia clínica completa	0,8421	0,4680	0,08190	40
Random Forest — sin familia TFG	0,7270	0,0691	0,01171	40
Random Forest — clínico/biológico sin observación	0,8823	0,5015	0,00651	40
Random Forest — proceso de observación	0,7434	0,0430	0,01601	40

Incertidumbre y robustez

Para el Random Forest de familia clínica completa, el bootstrap de 1.500 remuestras produjo un intervalo empírico del 95% de 0,7912 a 0,9436 para ROC-AUC, de 0,3433 a 0,6645 para

PR-AUC y de 0,00439 a 0,00812 para Brier. Las medianas fueron 0,8721, 0,5060 y 0,00609, respectivamente.

Tabla 8.

Incertidumbre del Random Forest mediante bootstrap

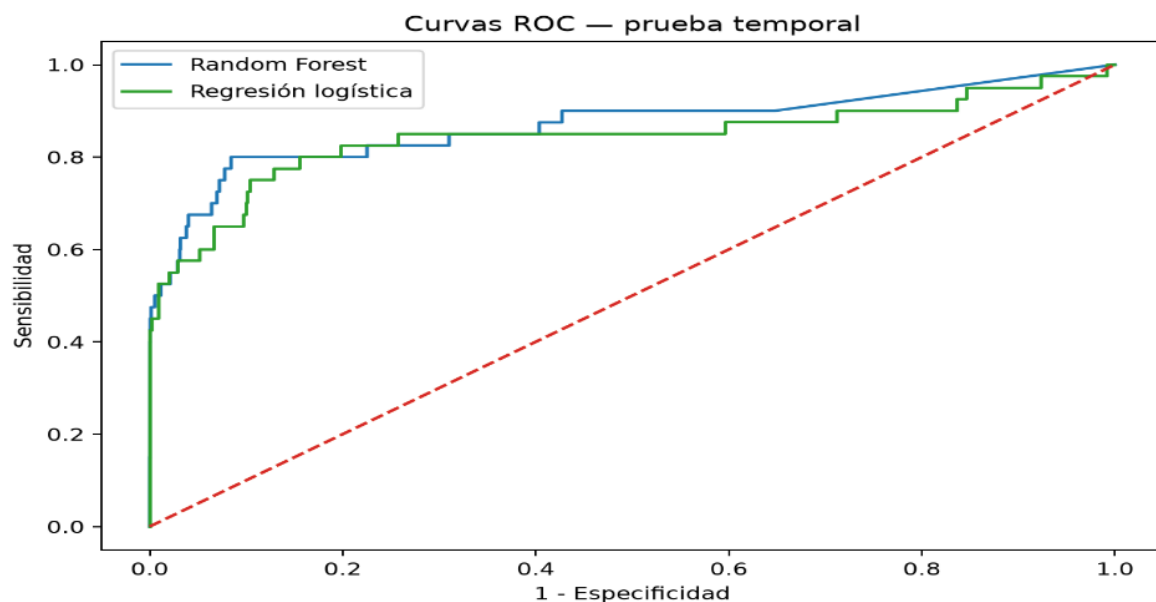
Métrica	Estimación	Bootstrap 95% (percentiles)	n
ROC-AUC	0,8689	0,7912–0,9436	1.500
PR-AUC	0,4956	0,3433–0,6645	1.500
Brier	0,00619	0,00439–0,00812	1.500

La robustez entre las semillas 7, 21, 42, 77 y 101 mostró ROC-AUC media 0,8680 (DE 0,0072), PR-AUC media 0,5060 (DE 0,0108) y Brier medio 0,006210 (DE 0,000069). La variación fue relativamente pequeña entre ejecuciones, aunque la incertidumbre continúa condicionada por el bajo número de eventos.

Curvas de discriminación

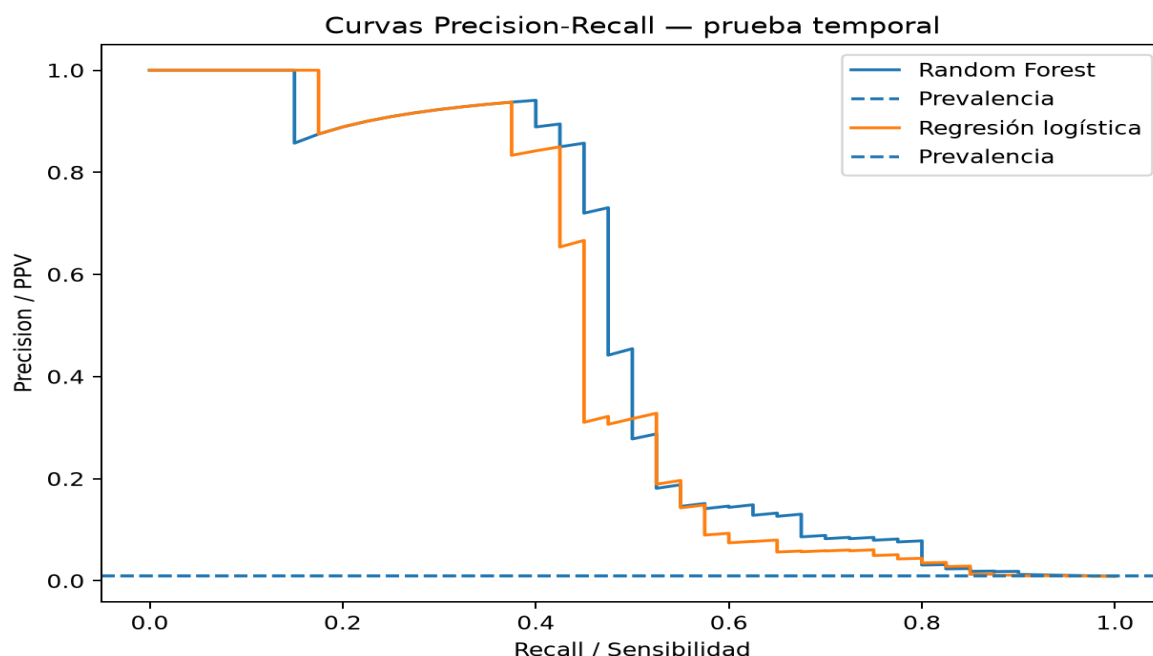
Figura 2.

Curva ROC de la validación temporal



Fuente: Elaboración propia mediante PALIC-PY-10 V12.1.0.

Figura 3.
Curva Precision-Recall de la validación temporal

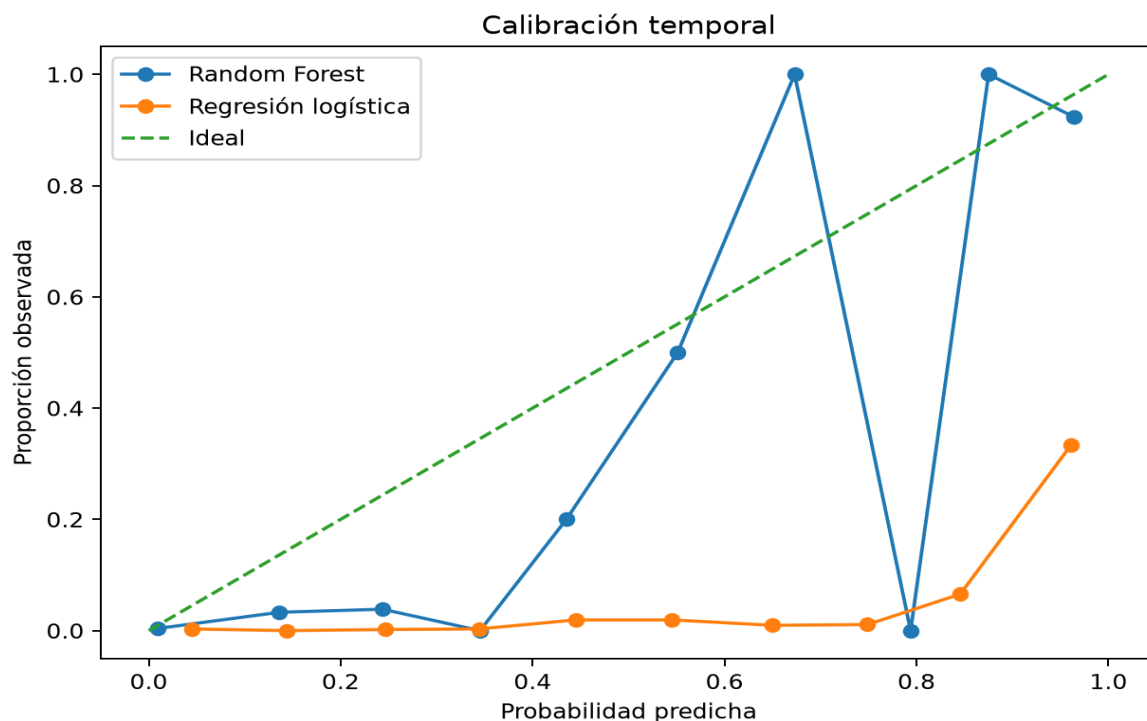


Fuente: Elaboración propia mediante PALIC-PY-10 V12.1.0.

Calibración

La calibración del Random Forest presentó intercepto -1,699 y pendiente 0,738. La pendiente inferior a 1 indica que las predicciones extremas tienden a requerir recalibración antes de interpretarse como probabilidades clínicas transportables. La incertidumbre en los estratos de mayor riesgo debe interpretarse con cautela debido al número reducido de eventos.

Figura 4.
Calibración temporal del modelo en el conjunto de prueba



Fuente: Elaboración propia mediante PALIC-PY-10 V12.1.0.

Importancia de variables

La evaluación de variables mostró que la señal predictiva estuvo fuertemente vinculada con la trayectoria renal histórica y la función renal disponible en el índice. La eliminación de la familia TFG/creatinina redujo el desempeño del Random Forest a ROC-AUC 0,7270 y PR-AUC 0,0691. El modelo clínico/biológico sin variables del proceso de observación alcanzó ROC-AUC 0,8823 y PR-AUC 0,5015, mientras que el proceso de observación por sí solo alcanzó ROC-AUC 0,7434 y PR-AUC 0,0430. Estas sensibilidades no prueban causalidad de vigilancia; muestran que la estructura de observación contiene señal parcial y que el desempeño no depende exclusivamente de ella.

Tabla 9.

Subgrupos considerados para evaluación exploratoria

Variable	Categoría	N
Sexo	F	3047
Sexo	M	1465
Región	CESAR	4038
Patología	HTA	2755
Patología	HTA+DM	1472
Grupo etario	45–59	1290
Grupo etario	60–74	2104
Grupo etario	≥75	875
Estadio ERC	G1	534
Estadio ERC	G2	2315
Estadio ERC	G3b	464

Comportamiento por subgrupos

En la prueba temporal se observaron diferencias exploratorias por sexo, región, patología, edad y estadio renal. Por ejemplo, el Random Forest obtuvo ROC-AUC 0,9195 en mujeres y 0,7911 en hombres; 0,9181 en HTA y 0,8402 en HTA+DM; y 0,9642 en el grupo de 45–59 años frente a 0,6854 en ≥75 años. En estadio renal, G1 mostró 0,9759, G2 0,5988 y G3b 0,7333. Estos resultados son exploratorios, están condicionados por el bajo número de eventos en varios subgrupos y no constituyen una evaluación de fairness ni validación externa.

Tabla 10.

Distribución de landmarks y eventos por categorías clínicas

Variable	Categoría	N	Eventos	Evento
Sexo	F	12024	71	0.590%
Sexo	M	5769	53	0.919%
Region	ATLANTICO	2720	7	0.257%
Region	CESAR	15073	117	0.776%
Patologia	DM	1252	5	0.399%
Patologia	HTA	10641	63	0.592%
Patologia	HTA+DM	5900	56	0.949%
Grupo_Etario	18-44	1087	5	0.460%
Grupo_Etario	45-59	5310	36	0.678%
Grupo_Etario	60-74	7994	51	0.638%
Grupo_Etario	75 o mas	3402	32	0.941%
EstadioERC_indice	G1	2182	61	2.796%
EstadioERC_indice	G2	9297	18	0.194%
EstadioERC_indice	G3a	4040	8	0.198%
EstadioERC_indice	G3b	1726	19	1.101%
EstadioERC_indice	G4	431	9	2.088%
EstadioERC_indice	G5	117	9	7.692%

Discusión

El principal hallazgo del estudio fue que una arquitectura longitudinal con control temporal explícito permitió construir un modelo con capacidad discriminativa relevante para ordenar el riesgo de un desenlace renal poco frecuente en una población real de programas de riesgo cardiovascular. El Random Forest alcanzó ROC-AUC 0,8689 y PR-AUC 0,4956 en la prueba temporal, con un Brier de 0,00619. Estos resultados deben interpretarse como evidencia de desempeño interno prometedor, no como evidencia de utilidad clínica ni como validación de un instrumento listo para implementación.

La comparación regional con Rojas et al. (2025) es especialmente relevante. Su trabajo mostró que es posible predecir la siguiente TFG en pacientes del Caribe colombiano mediante XGBoost y SHAP, con validación interna y externa. El presente estudio no pretende competir con esa finalidad predictiva. En lugar de estimar directamente la TFG futura, transforma la trayectoria renal en un desenlace binario sostenido. Esta diferencia modifica la pregunta clínica potencial: el objetivo no es pronosticar un número de TFG, sino identificar pacientes cuya trayectoria puede cruzar un umbral de deterioro relevante dentro del horizonte definido.

Un resultado metodológicamente importante fue la marcada diferencia entre el modelo clínico completo y la familia sin variables de TFG. El Random Forest completo presentó ROC-AUC 0,8689 y PR-AUC 0,4956, mientras que al excluir la familia TFG/creatinina el desempeño descendió a ROC-AUC 0,7270 y PR-AUC 0,0691. Esto confirma que el modelo es esencialmente un modelo de evolución renal y que su utilidad futura depende de la disponibilidad periódica de mediciones renales. El modelo clínico/biológico sin variables del proceso de observación mantuvo ROC-AUC 0,8823 y PR-AUC 0,5015, por lo que no debe atribuirse su rendimiento únicamente a la frecuencia de medición.

Implicaciones para investigación y gestión del riesgo

- El modelo puede servir como base para estudiar estratificación longitudinal del riesgo renal en programas de riesgo cardiovascular.
- La arquitectura point-in-time puede facilitar auditorías reproducibles y comparación entre instituciones con datos longitudinales.

- La baja prevalencia del evento hace indispensable evaluar PR-AUC, calibración y utilidad clínica antes de establecer umbrales operativos.
- La disponibilidad de TFG seriada es crítica; el modelo no debe presentarse como sustituto de la valoración clínica o de laboratorio.
- La validación externa y la evaluación prospectiva son requisitos para avanzar desde evidencia metodológica hacia evidencia de implementación.

Conclusiones

Se desarrolló un modelo longitudinal para identificar el riesgo de deterioro sostenido de la función renal en pacientes de programas de riesgo cardiovascular del Caribe colombiano, utilizando una cohorte de 18.326 landmarks, de los cuales 15.430 fueron evaluables para un horizonte de 365 días, con 124 eventos primarios.

El Random Forest mostró desempeño interno temporal prometedor, con ROC-AUC 0,8689, PR-AUC 0,4956 y Brier 0,00619. La robustez entre cinco semillas fue ROC-AUC $0,8680 \pm 0,0072$ y PR-AUC $0,5060 \pm 0,0108$. El bootstrap de 1.500 remuestras mostró una incertidumbre considerable, coherente con la baja frecuencia del evento.

La arquitectura PALIC-PY-10 V12.1.0 permitió documentar y auditar la temporalidad de los predictores, la construcción del landmark, el desenlace, la prevención de leakage, el split temporal canónico, la trazabilidad de predicciones, los hashes, el entorno y el self-check final. El resultado de integridad fue leakage READY y self-check PASS.

Los resultados no demuestran utilidad clínica, diagnóstico, causalidad ni validez externa. La aplicación futura requiere validación independiente, evaluación de calibración, análisis por subgrupos y estudios de utilidad clínica antes de cualquier implementación.

El trabajo aporta evidencia aplicada y una infraestructura metodológica propia para continuar una línea de investigación longitudinal en salud renal y riesgo cardiovascular en el Caribe colombiano. Su principal fortaleza es la trazabilidad de la cadena analítica; su principal condición para avanzar es la validación externa y prospectiva.

Referencias bibliográficas

- Díaz Lora, J. de J. (2026). Modelo analítico longitudinal aplicado a la identificación de riesgo de abandono y deterioro clínico en programas de riesgo cardiovascular en población del Caribe colombiano. *Revista Hambatu Science*, 1(2), 66–80. <https://doi.org/10.63862/rhs-v1n2-66-80-2026>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5–32. <https://doi.org/10.1023/A:1010933404324>
- Chan, L., et al. (2021). Derivation and validation of a machine learning risk score using biomarker and electronic patient data to predict progression of diabetic kidney disease. *Diabetologia*, 64, 1504–1515. <https://doi.org/10.1007/s00125-021-05444-0>
- Collins, G. S., Moons, K. G. M., Dhiman, P., Riley, R. D., Beam, A. L., van Calster, B., Ghassemi, M., Liu, X., Reitsma, J. B., van Smeden, M., et al. (2024). TRIPOD+AI statement: Updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ*, 385, e078378. <https://doi.org/10.1136/bmj-2023-078378>
- Kidney Disease: Improving Global Outcomes (KDIGO) CKD Work Group. (2024). KDIGO 2024 clinical practice guideline for the evaluation and management of chronic kidney disease. *Kidney International*, 105(Suppl. 4S), S117–S314. <https://doi.org/10.1016/j.kint.2023.10.018>
- Moons, K. G. M., Wolff, R. F., Riley, R. D., Whiting, P. F., Westwood, M., Collins, G. S., Reitsma, J. B., Kleijnen, J., & Mallett, S. (2019). PROBAST: A tool to assess risk of bias and applicability of prediction model studies. *Annals of Internal Medicine*, 170(1), 51–58. <https://doi.org/10.7326/M18-1376>
- Cuenta de Alto Costo. (2024). Situación de la enfermedad renal crónica, la hipertensión arterial y la diabetes mellitus en Colombia 2023. Fondo Colombiano de Enfermedades de Alto Costo.

Rojas, L. H., Pereira-Morales, A. J., Amador, W., Montenegro, A., Buelvas, W., & de la Espriella, V. (2025). Development and validation of interpretable machine learning models to predict glomerular filtration rate in chronic kidney disease Colombian patients. *Annals of Clinical Biochemistry*, 62(1), 57–66. <https://doi.org/10.1177/00045632241285528>

Tangri, N., et al. (2022). Development and external validation of a machine learning model for progression of chronic kidney disease. *Kidney International Reports*, 7(8), 1772–1781. <https://doi.org/10.1016/j.ekir.2022.05.004>

Tangri, N., Ferguson, T. W., Bamforth, R. J., Leon, S. J., Arnott, C., Mahaffey, K. W., Kotwal, S., Heerspink, H. J. L., Perkovic, V., Fletcher, R. A., & Neuen, B. L. (2024). Machine learning for prediction of chronic kidney disease progression: Validation of the Klinrisk model in the CANVAS Program and CREDENCE trial. *Diabetes, Obesity and Metabolism*, 26(8), 3371–3380. <https://doi.org/10.1111/dom.15678>

Declaraciones finales

Conflicto de intereses: Los autores declaran no tener ningún conflicto de interés relacionado con el presente trabajo.

Financiamiento: Esta investigación no recibió financiamiento externo de agencias públicas, comerciales o sin fines de lucro.

Agradecimiento: A mi querida esposa, **Marleny**, y a mi madre, **Etelvina**, por acompañarme y sostenerme con su apoyo constante en este camino de investigación. A **Lucas**, mi perro y compañero de vida durante 14 años, quien más que una mascota fue parte de mi familia. Aunque recientemente partió, su fidelidad, amor incondicional y compañía hicieron parte esencial de mi vida, durante el desarrollo de estos proyectos, sin saberlo, se convirtió en una fuente cotidiana de motivación para continuar.

Uso de inteligencia artificial generativa: NA.

Nota editorial: El manuscrito no reproduce una publicación anterior. Aunque utiliza la misma fuente poblacional general de base de la línea de investigación del Artículo 1, el presente trabajo

formula una pregunta diferente, reconstruye una cohorte renal longitudinal específica, utiliza un desenlace operacional distinto, excluye las variables agregadas del Artículo 1 y emplea una arquitectura temporal y un pipeline PALIC-PY-10 V12.1.0 propios de este estudio. Los datos clínicos de origen no son públicos por contener información derivada de atención en salud. La implementación metodológica y los artefactos de auditoría pueden conservarse como evidencia reproducible del proyecto, sujetos a las restricciones aplicables a la fuente de datos.